FINAL REPORT

US AFOSR Grant FA9550-07-1-0400

Title: Blind Spread-Spectrum Steganalysis via Iterative Techniques

Principal Investigators:
Prof. Stella N Batalama, Assoc. Dean for Research,
and Prof. Dimitris A Pados
Department of Electrical Engineering
University at Buffalo, The State University of New York
{batalama, pados}@buffalo.edu

Date: June 17, 2010

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
17-06-2010

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*
April 1, 2007 - November 30, 2009

4. TITLE AND SUBTITLE
Blind Spread-Spectrum Steganalysis via Iterative Techniques

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-07-1-0400

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Stella N Batalama and Dimitris A Pados

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The Research Foundation of SUNY
University at Buffalo
402 Crofts Hall
Buffalo, NY 14260

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/RSL
875 North Randolph Street
Arlington, VA 22203-1954

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR/RSL

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-SR-AR-TR-10-0265

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION A: APPROVED FOR PUBLIC RELEASE

13. SUPPLEMENTARY NOTES
N/A

14. ABSTRACT

We developed theory and methods for optimal digital data hiding in arbitrary transform domains of digital hosts (images, video, audio). Our optimality criteria are: Host distortion, recovery error rate, and the Shannon capacity of the covert channel. Additionally, we introduced for the first time the concept of multiuser/multi-signature steganography. Then, we developed counter-measures to (optimal multiuser) steganography in the form of active (message extraction) and passive (stego/non-stego decision) steganalysis. We concluded that optimal data hiding, as described in the report, offers vast improvement in recovery error rate/Shannon capacity versus medium distortion and enables highly effective multi-signature embedding (different -potentially- hidden messages for different POCs along the chain of command etc.). In the context of active steganalysis, the developed M-IGLS hidden message extraction algorithm can destroy conventional SS steganography. However, our own optimal SS embedding is resistant to M-IGLS steganalysis attacks, especially for small hidden messages. Our new passive (binary hypothesis testing) steganalysis procedure offers close to 95% identification success rate at about 1% false alarm rate when used on hosts with conventionally embedded messages.

15. SUBJECT TERMS
Steganography, steganalysis, data hiding, embedding, watermarking, authentication, covert communications.

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
U	U	U	UU	11

19a. NAME OF RESPONSIBLE PERSON
Stella N Batalama

19b. TELEPHONE NUMBER *(Include area code)*
(716) 645-1147

Reset

I. Overview

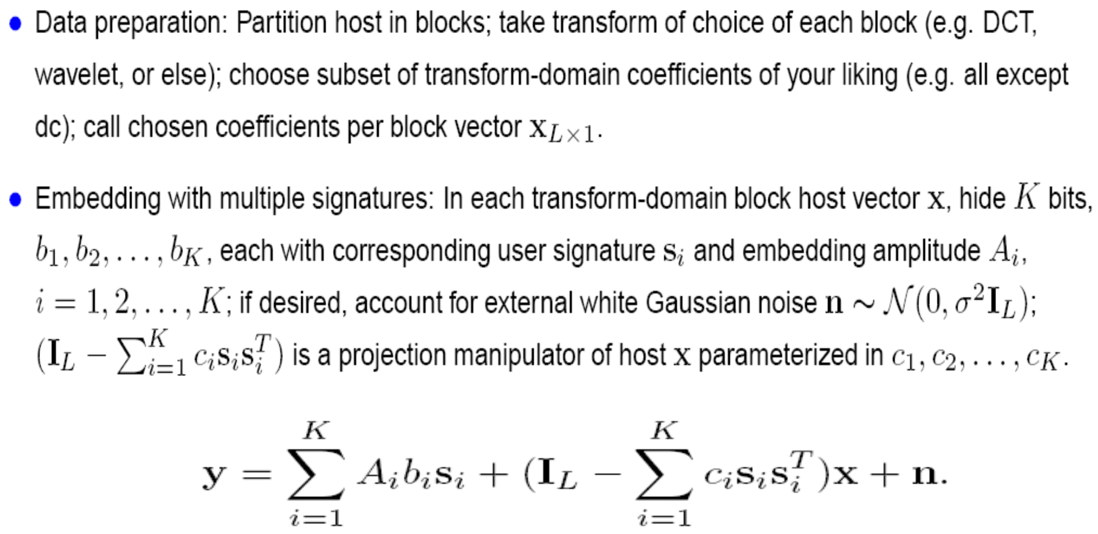
We begin with an outline of the research effort.

We developed theory and methods for optimal digital data hiding in arbitrary transform domains of digital hosts (images, video, audio). Our optimality criteria are mean-square host distortion, recovery error rate, and the Shannon capacity of the covert channel. Additionally, we introduced for the first time the concept of multiuser/multi-signature steganography.

Finally, we developed new counter-measures to (optimal multiuser) steganography in the form of active (message extraction) and passive (stego/non-stego decision) steganalysis.

II. Research Breakthrough: Optimal Multiuser Embedding

The following two steps describe in a most concise manner the developed optimal embedding (data hiding) procedure.



- Data preparation: Partition host in blocks; take transform of choice of each block (e.g. DCT, wavelet, or else); choose subset of transform-domain coefficients of your liking (e.g. all except dc); call chosen coefficients per block vector $\mathbf{x}_{L\times 1}$.
- Embedding with multiple signatures: In each transform-domain block host vector $\mathbf{x}$, hide K bits, $b_1, b_2, \ldots, b_K$, each with corresponding user signature $\mathbf{s}_i$ and embedding amplitude A_i, $i = 1, 2, \ldots, K$; if desired, account for external white Gaussian noise $\mathbf{n} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_L)$; $(\mathbf{I}_L - \sum_{i=1}^{K} c_i \mathbf{s}_i \mathbf{s}_i^T)$ is a projection manipulator of host $\mathbf{x}$ parameterized in $c_1, c_2, \ldots, c_K$.

$$\mathbf{y} = \sum_{i=1}^{K} A_i b_i \mathbf{s}_i + (\mathbf{I}_L - \sum_{i=1}^{K} c_i \mathbf{s}_i \mathbf{s}_i^T)\mathbf{x} + \mathbf{n}.$$

The optimal embedding signatures and scalar parameters to be used in the above equation are tabulated below.

Optimal $(\mathbf{s}_i^{\mathrm{opt}}, c_i^{\mathrm{opt}})$ pairs, $i = 1, \ldots, K$:

$$\mathbf{s}_i^{\mathrm{opt}} = \mathbf{q}_{L-i+1}$$

$$c_i^{\mathrm{opt}} = \frac{\lambda_{L-i+1} + \sigma_n^2 \mathcal{D}_i - \sqrt{(\lambda_{L-i+1} + \sigma_n^2 + \mathcal{D}_i)^2 - 4\lambda_{L-i+1}\mathcal{D}_i}}{2\lambda_{L-i+1}}$$

where $q_1, \ldots, q_L$ are eigenvectors of $\mathrm{R}_{\mathrm{x}} = \mathrm{E}\{\mathrm{x}\mathrm{x}^T\}$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_L$ and $\mathcal{D}_i = A_i^2 + c_i \mathbf{s}_i^T \mathrm{R}_{\mathrm{x}} \mathbf{s}_i, i = 1, \ldots, K$.

These assignments complete the description of optimal multiuser steganography.

III. Steganography Experimental Studies

Below, we present an example where the optimal procedure of Section II is applied. Figure 1 shows the original 256 x 256 gray-scale "Baboon" image.

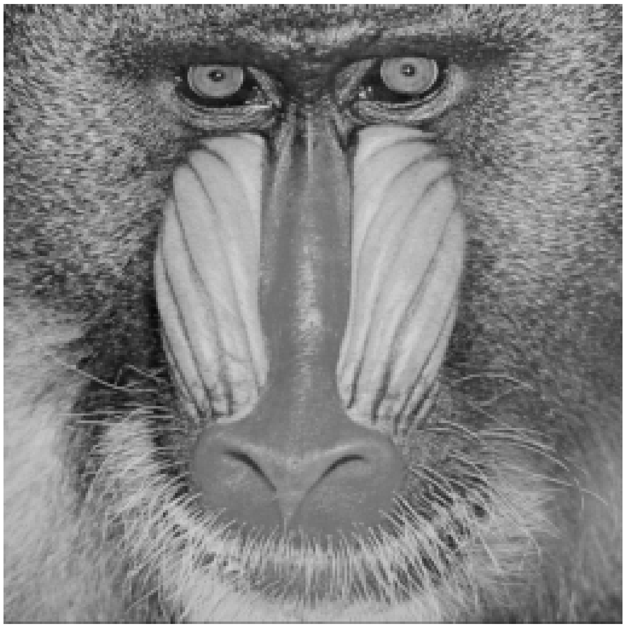


Figure 1

Figure 2 shows the same image after optimal embedding of K=15 messages of size 1Kbit each, equal per-message distortion, total distortion 31.8dB, and additive white Gaussian noise -for the sake of generality- of 3dB.

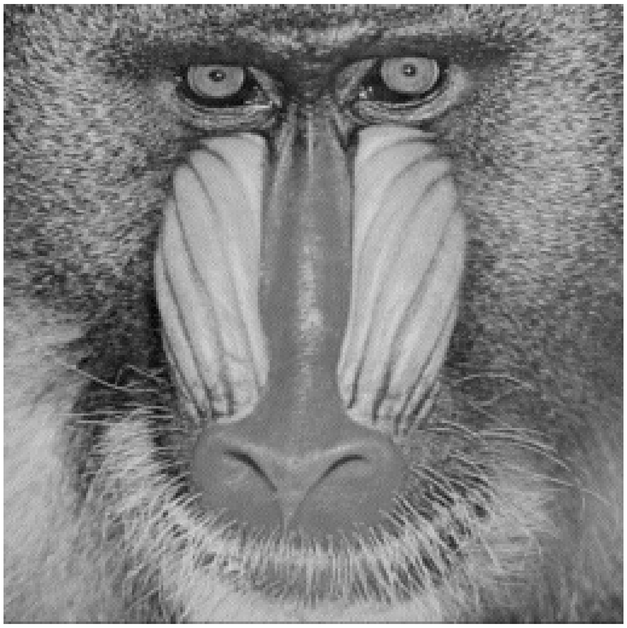


Figure 2

Figure 3, below, shows the bit-error-rate (BER) versus host for the 15 messages hidden in Fig. 2.

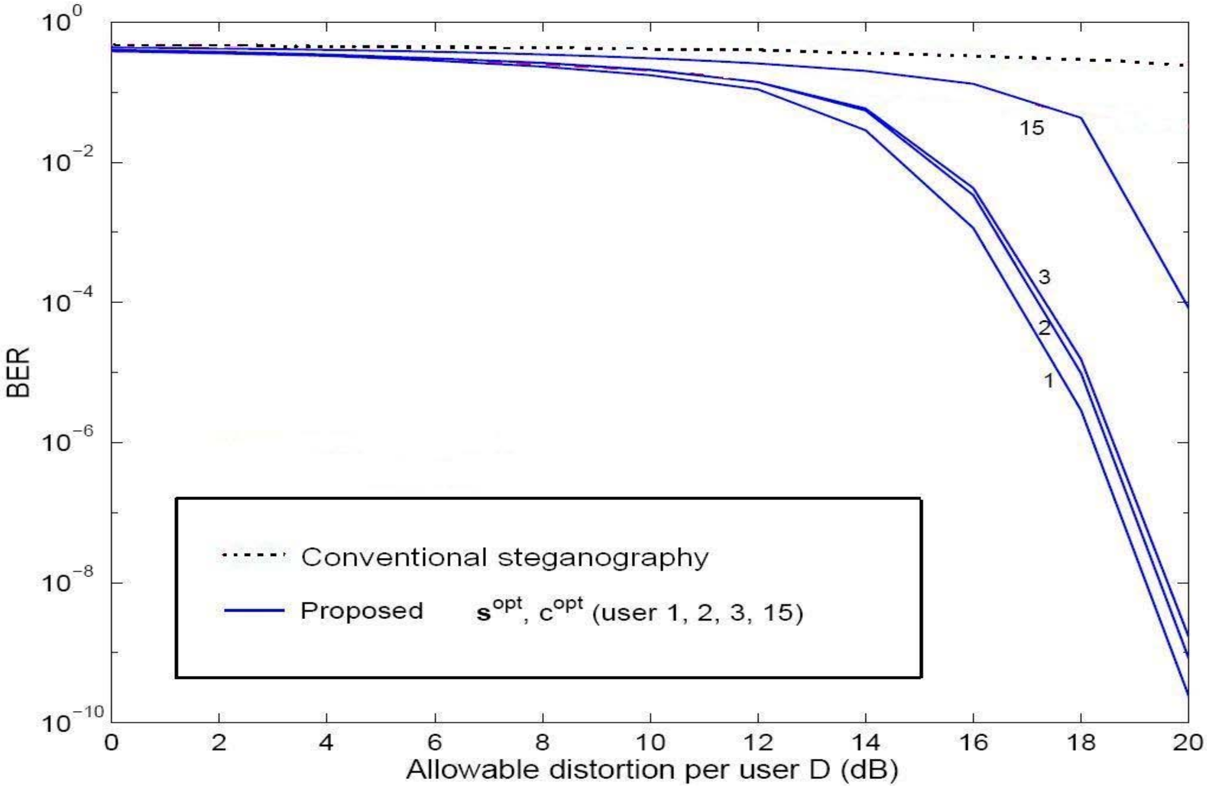


Figure 3

In Figure 4, we present the sum-capacity of the covert channel as a function of the total distortion (K=15 messages). The presented result is the average over the whole USC-SIPI image database.

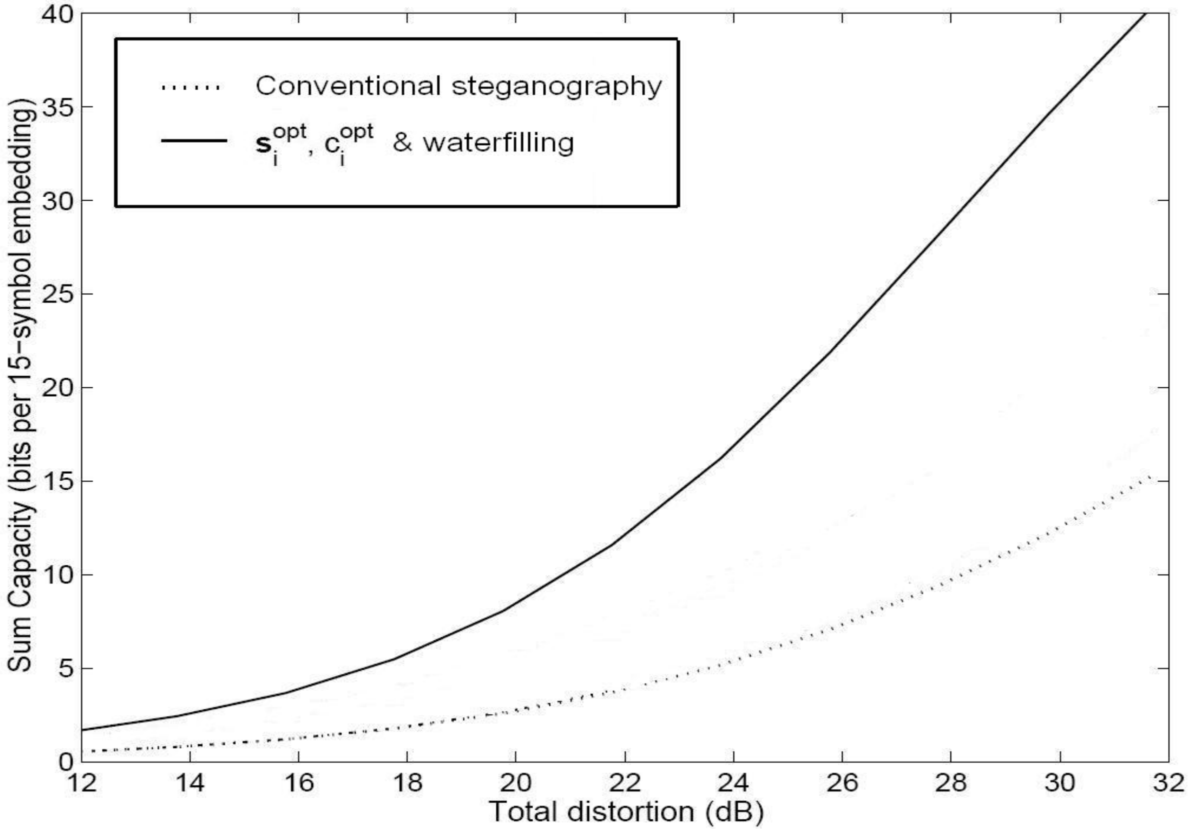
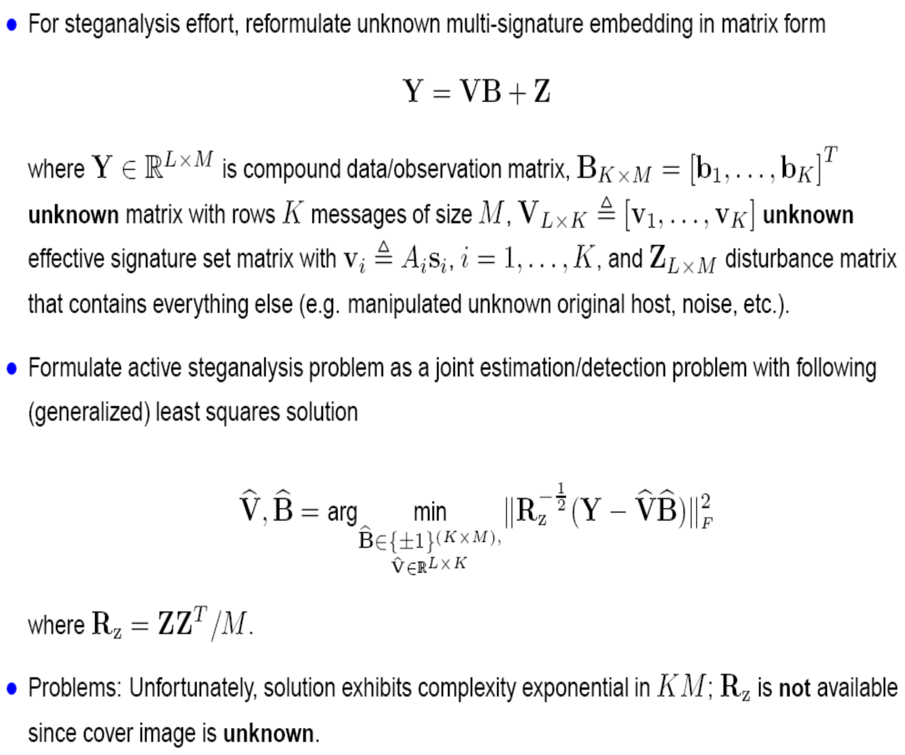


Figure 4

IV. Research Breakthrough: Active Multi-signature Steganalysis

In the following, we present the steps of the novel multi-signature iterative-generalized-least-squares (M-IGLS) procedure that we developed which, as demonstrated later on, enables effective recovery of messages hidden by conventional spread-spectrum embedding means even when the embedding signatures are completely unknown.

We begin with a careful formulation of the problem.



- For steganalysis effort, reformulate unknown multi-signature embedding in matrix form

$$\mathrm{Y} = \mathrm{V}\mathrm{B} + \mathrm{Z}$$

where $\mathrm{Y} \in \mathbb{R}^{L\times M}$ is compound data/observation matrix, $\mathrm{B}_{K\times M} = [\mathbf{b}_1, \ldots, \mathbf{b}_K]^T$ **unknown** matrix with rows K messages of size M, $\mathrm{V}_{L\times K} \triangleq [\mathrm{v}_1, \ldots, \mathrm{v}_K]$ **unknown** effective signature set matrix with $\mathrm{v}_i \triangleq A_i \mathrm{s}_i$, $i = 1, \ldots, K$, and $\mathrm{Z}_{L\times M}$ disturbance matrix that contains everything else (e.g. manipulated unknown original host, noise, etc.).

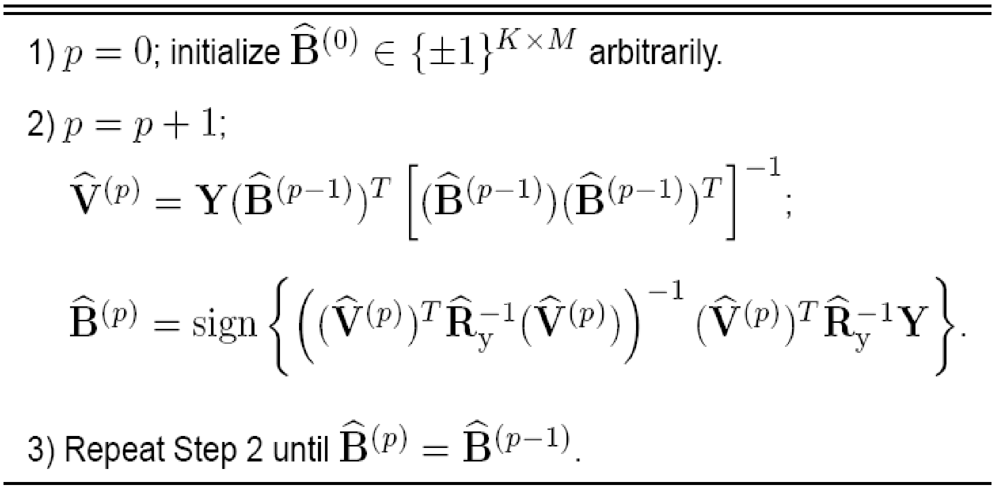
- Formulate active steganalysis problem as a joint estimation/detection problem with following (generalized) least squares solution

$$\widehat{\mathrm{V}}, \widehat{\mathrm{B}} = \arg \min_{\substack{\widehat{\mathrm{B}} \in \{\pm 1\}^{(K\times M)}, \\ \widehat{\mathrm{V}} \in \mathbb{R}^{L\times K}}} \|\mathrm{R}_{\mathrm{z}}^{-\frac{1}{2}}(\mathrm{Y} - \widehat{\mathrm{V}}\widehat{\mathrm{B}})\|_F^2$$

where $\mathrm{R}_{\mathrm{z}} = \mathrm{Z}\mathrm{Z}^T/M$.

- Problems: Unfortunately, solution exhibits complexity exponential in KM; R_{z} is **not** available since cover image is **unknown**.

We can, however, overcome the identified problems effectively by first replacing in the optimization formula $\mathbf{R_z}$ by $\widehat{\mathbf{R}}_{\mathbf{y}} = \mathbf{Y}\mathbf{Y}^T/M$ and then by running the following procedure that was theoretically derived applying iterative mean-square principles.



```
1) $p = 0$; initialize $\widehat{\mathrm{B}}^{(0)} \in \{\pm 1\}^{K\times M}$ arbitrarily.
2) $p = p + 1$;
    $\widehat{\mathbf{V}}^{(p)} = \mathbf{Y}(\widehat{\mathbf{B}}^{(p-1)})^T \left[(\widehat{\mathbf{B}}^{(p-1)})(\widehat{\mathbf{B}}^{(p-1)})^T\right]^{-1}$;
    $\widehat{\mathbf{B}}^{(p)} = \mathrm{sign}\left\{\left((\widehat{\mathbf{V}}^{(p)})^T \widehat{\mathbf{R}}_{\mathrm{y}}^{-1}(\widehat{\mathbf{V}}^{(p)})\right)^{-1} (\widehat{\mathbf{V}}^{(p)})^T \widehat{\mathbf{R}}_{\mathrm{y}}^{-1}\mathbf{Y}\right\}$.
3) Repeat Step 2 until $\widehat{\mathrm{B}}^{(p)} = \widehat{\mathrm{B}}^{(p-1)}$.
```

V. Steganalysis Experimental Studies

Here, we examine the performance of the developed active steganalysis procedure of Section IV when, first, unknown data hiding was carried out by conventional spread-spectrum embedding means and, next, when embedding was done by the optimal procedure of Section II.

Figure 5 plots the average bit-error-rate versus per message distortion when K=4 messages of length 1Kbit are embedded by conventional spread-spectrum means in the 256 x 256 Baboon image of Figure 1 together with 3dB additive white Gaussian noise. Our steganalysis algorithm performance (black line) shows that the intended recipient has (little) advantage only when they use for recovery the optimal minimum-mean-square-error (MMSE) filter and not just the embedding signature. Practically, our steganalysis algorithm renders such steganography unusable/useless.

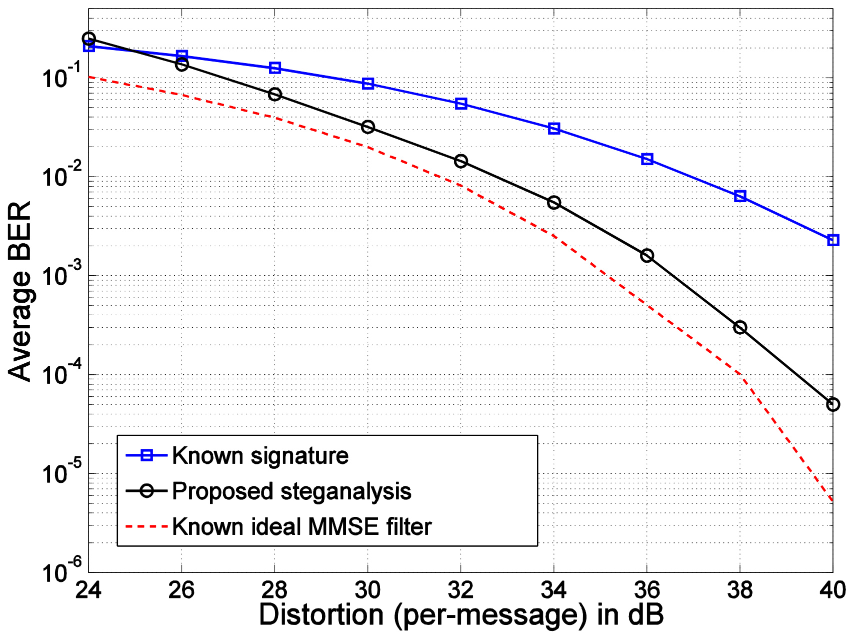


Figure 5

In Figure 6, we carry out the exact same study but for optimal data hiding as described in Section II. We observe that at the 19 to 20dB -and thereafter- range a gap opens up between our steganalysis BER (black line) and the BER of the intended recipient (blue line), which opens up a window of opportunity for effective steganography.

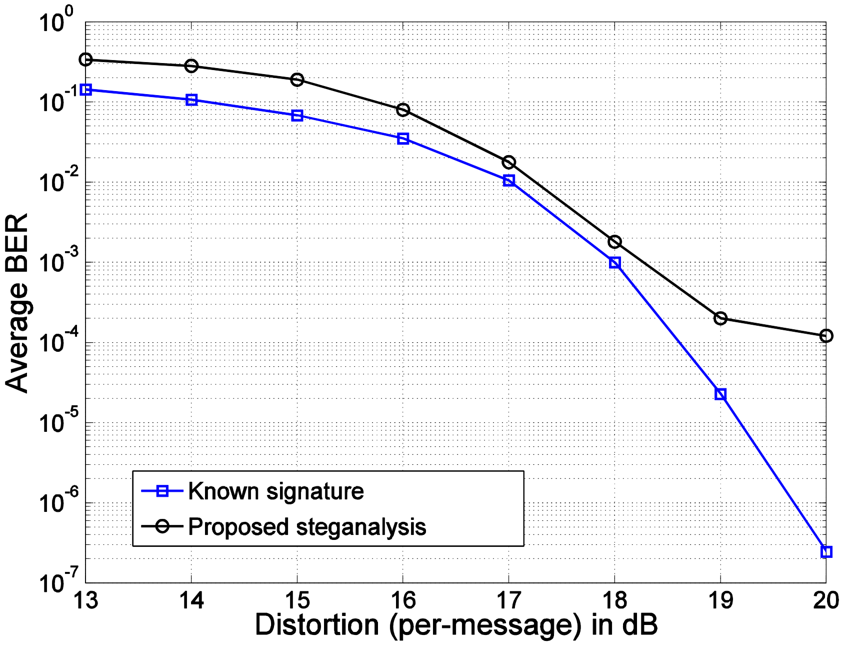


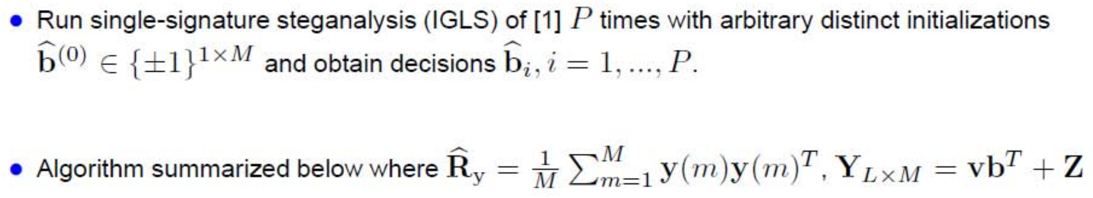
Figure 6

IV. Research Breakthrough: Passive Multisignature Steganalysis

In our language, passive steganalysis is the problem of deciding in favor of either presence or absence of a (multi-signature) spread-spectrum hidden message(s) in a given digital medium. It is, therefore, a binary hypothesis testing problem. Passive steganalysis is envisioned as a rapid high-volume scanning technology that identifies and sets aside for further scrutiny suspicious media. With this understanding, we set our own passive steganalysis requirements as follows. Our algorithm must be of low complexity (for rapid scanning operation), image/medium independent (broad applicability without modifications), and unsupervised (we should not be expecting embedding examples by our foes).

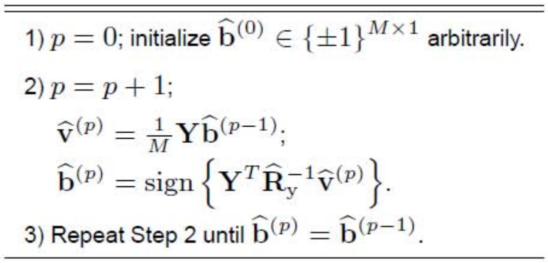
The developed algorithmic procedure (low-complexity, medium-independent, and unsupervised) is as follows.

STEP 1:



- Run single-signature steganalysis (IGLS) of [1] P times with arbitrary distinct initializations $\widehat{\mathbf{b}}^{(0)} \in \{\pm 1\}^{1\times M}$ and obtain decisions $\widehat{\mathbf{b}}_i$, $i = 1, ..., P$.

- Algorithm summarized below where $\widehat{\mathbf{R}}_{\mathrm{y}} = \frac{1}{M}\sum_{m=1}^{M} \mathbf{y}(m)\mathbf{y}(m)^T$, $\mathbf{Y}_{L\times M} = \mathbf{v}\mathbf{b}^T + \mathbf{Z}$

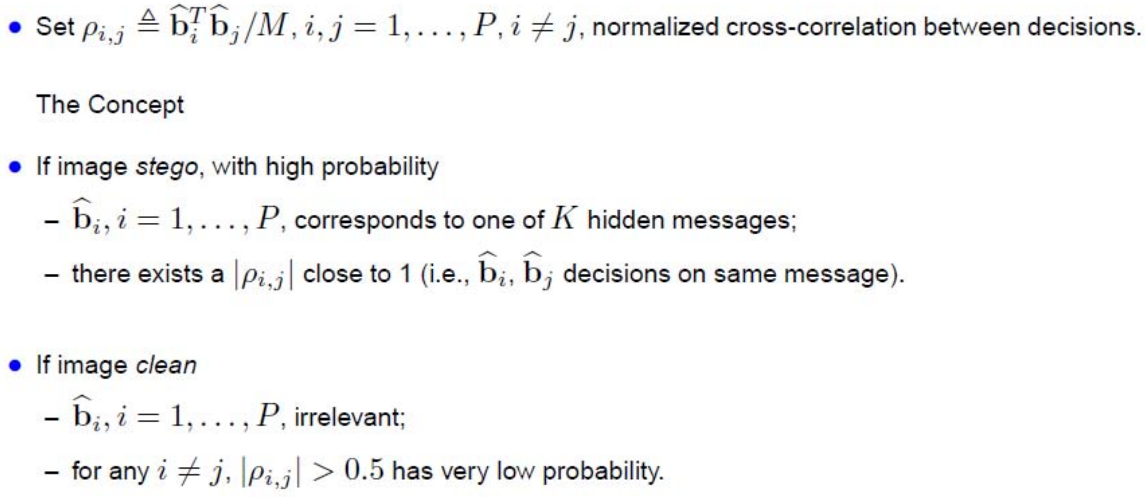


```
1) $p = 0$; initialize $\widehat{\mathbf{b}}^{(0)} \in \{\pm 1\}^{M\times 1}$ arbitrarily.
2) $p = p + 1$;
    $\widehat{\mathbf{v}}^{(p)} = \frac{1}{M}\mathbf{Y}\widehat{\mathbf{b}}^{(p-1)}$;
    $\widehat{\mathbf{b}}^{(p)} = \mathrm{sign}\left\{\mathbf{Y}^T\widehat{\mathbf{R}}_{\mathrm{y}}^{-1}\widehat{\mathbf{v}}^{(p)}\right\}$.
3) Repeat Step 2 until $\widehat{\mathbf{b}}^{(p)} = \widehat{\mathbf{b}}^{(p-1)}$.
```

[1] M. Gkizeli, D. A. Pados, S. N. Batalama, and M. J. Medley, " Blind iterative recovery of spread-spectrum steganographic messages," in *Proc. IEEE Intern. Conf. Image Proc.*, Genova, Italy, Sept. 2005, vol. 2, pp. 11-14.

Then, correlate among obtained decisions as follows.

STEP 2:



- Set $\rho_{i,j} \triangleq \widehat{\mathbf{b}}_i^T\widehat{\mathbf{b}}_j/M$, $i, j = 1, \ldots, P$, $i \neq j$, normalized cross-correlation between decisions.

  The Concept

- If image *stego*, with high probability
  - $\widehat{\mathbf{b}}_i$, $i = 1, \ldots, P$, corresponds to one of K hidden messages;
  - there exists a $|\rho_{i,j}|$ close to 1 (i.e., $\widehat{\mathbf{b}}_i$, $\widehat{\mathbf{b}}_j$ decisions on same message).

- If image *clean*
  - $\widehat{\mathbf{b}}_i$, $i = 1, \ldots, P$, irrelevant;
  - for any $i \neq j$, $|\rho_{i,j}| > 0.5$ has very low probability.

Here is now the complete passive steganalysis algorithm.

```
Stego=0; $i = 0$.
While $i \leq P$
    $i = i + 1$.
    Execute Gkizeli-Pados-Batalama-Medley routine [1] and obtain decision $\widehat{\mathbf{b}}_i$.
    If $|\rho_{i,j}| > \gamma$ for any $1 \leq j < i$
        Stego=1; $i = P + 1$.
    End
End
```

- Threshold γ usually chosen in $[0.5, 0.9]$ range; larger γ induces lower false alarm rate P_{FA} but higher probability of miss P_M and vice versa;
- $P = 30$ to 200.

The algorithm shows exceptional promise against conventional spread-spectrum steganography as demonstrated in our experiments. Figure 7(a) and its zoom-in in (b) show probability of correct identification versus false alarm on a dataset of about 1,500 images [3], [4] and compare against the recent feature extraction algorithm in [2].

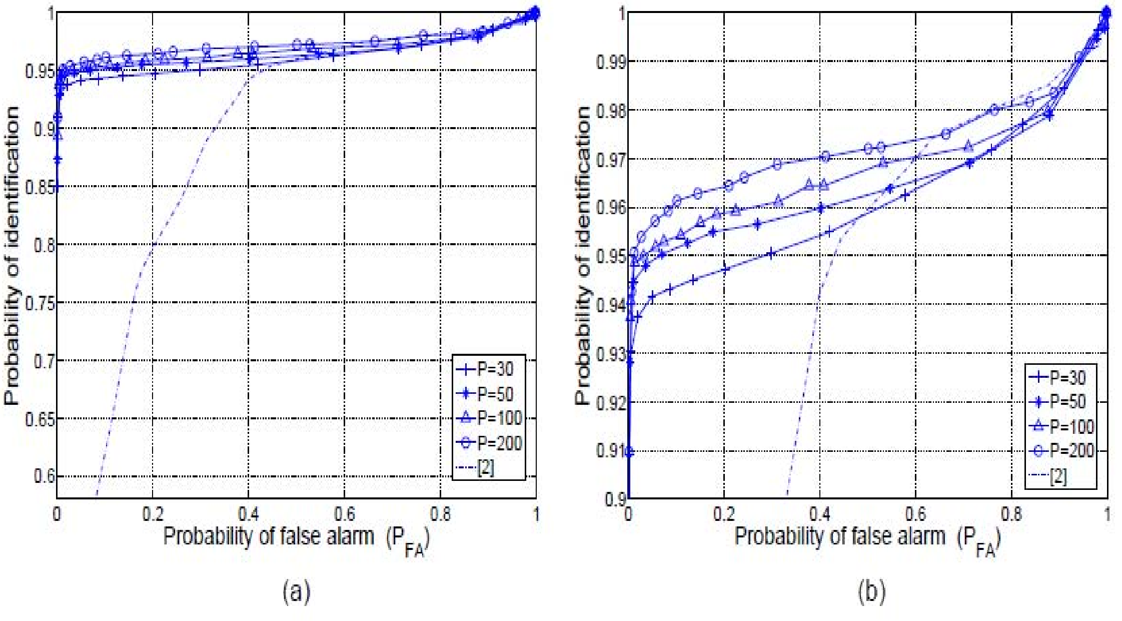


Figure 7

[2] Y. Wang and P. Moulin, "Optimized feature extraction for learning-based image steganalysis," *IEEE Trans. Inform. Forensics and Security*, vol. 2, pp. 31-45 , Mar. 2007.

[3] USC-SIPI Image Database, [Online]. Available: http://sipi.usc.edu/ database/database.cgi?volume=misc

[4] UCID-Uncompressed Colour Image Database, [Online]. Available: http://vision.cs.aston.ac.uk/datasets/UCID/ucid.html

VII. Concluding Remarks

Optimal data hiding, as described in this grant report, offers vast improvement in recovery error rate/Shannon capacity versus distortion and enables highly effective multi-signature embedding (different -potentially- hidden messages for different points of contact along the chain of command etc.).

Our developed active steganalysis M-IGLS hidden message extraction algorithm can destroy conventional SS steganography. However, our optimal embedding scheme is resistant to M-IGLS steganalysis attacks, especially for small hidden messages.

Our new passive (binary hypothesis testing) steganalysis procedure offers close to 95% identification success rate at about 1% false alarm when used on hosts with conventionally spread-spectrum embedded messages. We have not done tests yet on host with optimally embedded messages.

Our suggested plan for continued research is as follows.
Optimal steganography: Study and analysis of transforms, host partitions, multi-signature assignments, variable-length signature optimization.
Active steganalysis: Research on the effective recovery of relatively small messages embedded with own optimal scheme.
Passive steganalysis: Algorithmic tests and modifications against optimal embedding.
Video steganography and steganalysis: Pioneer this new research area with uncompressed (raw) and compressed video.